

IV. ИЗМЕНЧИВОСТЬ ГЕНОФОНДА ЧЕЛОВЕКА И МЕДИЦИНСКАЯ ГЕНЕТИКА

ПРОСТРАНСТВЕННАЯ ИЗМЕНЧИВОСТЬ ГЕНОФОНДА ЧЕЛОВЕКА: ГЕНОГЕОГРАФИЯ И ОТБОР*

Е. В. Балановская, С. Д. Нурбаев

Медико-генетический научный центр РАМН, Москва

Spatial variability of the human gene pool: gene geography and natural selection

E. V. Balanovskaya, S. D. Nurbaev

Medical Genetic Science Center of RAMS, Moscow

Methodological aspects of the detection of selection effects by the investigation of geographical variability of the gene pool are considered. Four aspects are analyzed in application to the total human gene pool.

1. Computer gene geography. Interpolational gene geographical maps can be created for single allele, for complexes of genes and for the gene pool as a whole. The correlation between gene geographical maps and environmental factors allows to suggest the selection hypothesis.

2. Selective structure of the gene pool. If interpopulation variability (F_{ST}) of some allele differs significantly from selectively neutral F_{ST} -level, it can suggest that this allele is under selection. The computer technology for the estimation of a type and intensity of the selection is designed (taking into account the real β -distribution of F_{ST} and modeling by means of numeral resampling).

3. Complex analysis of the gene pool. Being used in complex the technology of gene geography and that of the selective structure yield the most reliable results and allow to perform many-sided analysis of selection effects, to correct and to check hypotheses.

4. The reliability of models. For the first time, the quantitative estimation of the reliability of gene geographical maps (for each point of an area analyzed) is presented on the base of the theory of reliability. As a result, the selection hypothesis is suggested on the base of investigation of those parts of area only, which are reliably provided by original information.

Цель данного сообщения чисто методическая: рассмотреть те способы изучения пространственной изменчивости генофонда, которые позволяют оценить эффекты отбора, исследовать систему «ген–популяция–среда». В качестве объекта изучения выступает генофонд человека на различных уровнях популяционной структуры — от элементарных популяций до ойкумены в целом. Однако все изложенные методы и технологии применимы почти к любым живым организмам.

Проследить эволюцию популяций во времени сложно. Но есть иной путь исследования — через генетическое многообразие современных популяций. Многоликость современного населения является итогом и материальным воплощением всего хода их эволю-

* Работа выполнена по гранту РФФИ и ГНТП «Приоритетные направления генетики».

© Е. В. Балановская, С. Д. Нурбаев, 1999

ции, т. е. разверткой в пространстве процесса эволюции, протекающего во времени. Этот принцип эргодичности ярко выражен Э. Реклю: «География по отношению к человеку есть не что иное, как История в пространстве, подобно тому, как История является Географией во времени» [22, с. 1]. Такая «взаимозаменяемость» времени и пространства и позволяет изучать микроэволюцию популяций через их географическую изменчивость.

Из огромного массива данных о генетических различиях популяций в пределах ойкумены рассмотрим лишь имеющие прямое отношение к отбору. Сразу подчеркнем, что во всех случаях рассматривается *только коренное* население каждого региона мира.

Геногеографические карты

Геногеография включает в себя множество методов анализа изменчивости генетических признаков в пространстве. Однако наиболее универсальным и многогранным инструментом анализа являются картографические модели. Геногеографические карты — чрезвычайно чувствительный инструмент с большой разрешающей силой. Например, они позволили выявить следы летописных славянских племен в антропологическом облике современного русского населения [24], в то время как стандартными методами антропологии фиксировался лишь финно-угорский субстрат (тот пласт коренного населения — мурома, мешера и т. д., который был ассимилирован славянской культурой).

Процедура построения, анализа и чтения компьютерных геногеографических карт подробно изложена нами раньше [2–5]. Создается компьютерная картографическая основа (территории, акватории, реки, границы), на которую наносятся генетически изученные популяции. Далее накладывается как бы «листочек в клетку» — густая равномерная сеть. Для каждого узла сетки (их обычно несколько тысяч) с помощью интерполяционной процедуры рассчитывается значение признака: с учетом всех популяций в пределах заданного радиуса. Такая операция проводится независимо для каждого узла сетки. В результате получаем равномерную численную матрицу — цифровую модель (ЦМ) карты. Далее с ЦМ (как с обычными матрицами) можно проводить любые преобразования, получая количественные оценки связей и закономерностей. При этом принципиально важно, что карта становится не иллюстрацией, а методом количественного анализа пространственной изменчивости генофонда.

Чтобы перейти от матрицы к карте, объединяем значения матрицы в равномерные интервалы, указанные в легенде. Карта читается как обычная географическая карта — от минимума значений к максимуму. На цветных картах низкие значения признака обозначаются синим (как морские впадины), средние значения — зелеными равнины, высокие значения — коричневыми горы. На черно-белых картах интенсивность окраски соответствует величине признака: низкие значения — самые светлые, наиболее высокие — самые темные.

Чтобы выявить основные закономерности в изменчивости отдельного гена (аллеля) или генофонда в целом, мы вновь обращаемся не к интервалам, а к точным значениям ЦМ — именно с ней мы производим любые преобразования. Например, были построены карты 100 аллелей 30 локусов иммунологического и биохимического полиморфизма, наиболее изученных в Северной Евразии (в границах бывшего СССР). По их цифровым матрицам рассчитаны главные компоненты и вновь картированы.

Карты первых двух главных компонент отражают основную часть изменчивости различных аллелей, т. е. основные закономерности пространственной изменчивости генофонда в целом. Карта первой главной компоненты демонстрирует тренд, который пронизывает весь генофонд в направлении «запад–восток». И именно таково основное направле-

ние исторического процесса в этой части Евразии. Поэтому эту карту мы объясняем обычно с точки зрения истории — на протяжении всего времени развития населения этого региона шло взаимопроникновение двух генофондов: европейского и азиатского [12, 26]. Такая интерпретация была недавно подтверждена самой необычной нашей картой [16] — картой материальной культуры палеолита (период 26–16 тыс. лет назад). Картографировалось распространение всех основных элементов палеолитической культуры — характеристики орудий труда, жилища, орнамента, объектов охоты. На основе этих карт рассчитаны карты главных компонент культуры палеолита. Карта первой компоненты выявляет наиболее общие черты пространственной изменчивости *культуры*. Примерно по 70-му меридиану проходит резкая граница между двумя крупнейшими палеолитическими провинциями — Европейской и Сибирской. Поражает удивительное сходство карты этой с картой первой компоненты *современного генофонда* — лишь граница между европейским и сибирским генофондами становится широкой и размытой. Двадцать тысяч лет истории населения Евразии не стерли различия между этими двумя провинциями: генофонд хранит память о своей истории.

Карта второй по значимости главной компоненты генофонда воспроизводит картину широтной зональности природной среды в направлении «юг–север». Поэтому если карту первой компоненты мы соотносим с *Историей*, то карту второй компоненты можем соотносить со второй силой, управляющей эволюцией генов — силой *Природы*, т. е. с влиянием среды и отбора. Можно и количественно оценить соотношение этих двух эволюционных сил — по доле дисперсии: История оказывается почти в три раза мощнее Природы.

Однако подобная интерпретация карты — по ассоциации — всегда требует особой осторожности, поскольку количество объяснений прямо зависит от богатства ассоциативных связей интерпретатора. Например, частично изменчивость второй компоненты можно объяснить не только влиянием среды, но и историей населения: экспансией степного кочевого мира в таежную зону Евразии. Именно поэтому необходим комплексный анализ генофонда — *контроль гипотезы отбора со стороны независимых методов*. Рассмотрим один из таких методов — изучения эффектов отбора через пространственную изменчивость генофонда.

Структура отбора

Схема анализа такова: в пространственной изменчивости генов можно выделить две составляющих — селективно-нейтральную и селективно-значимую.

Селективно-нейтральная изменчивость обусловлена двумя факторами микроэволюции — миграциями и дрейфом генов. Они действуют безотносительно к функции гена и связаны только с демографией (характер брачных связей и генетически эффективный размер популяции), общей для всех генов. Именно поэтому селективно-нейтральный процесс вбирает в себя все общее в судьбах генов и отражает историю генофонда в целом. Он является той общей магистралью эволюции, которая задается миграциями и дрейфом генов. Определить селективно-нейтральный уровень изменчивости популяций можно, например, в среднем по совокупности независимых локусов [1, 6–9, 23, 25, 28, 30] с помощью F_{ST} -статистики Райта или G_{ST} -статистики Нея или, например, шенноновской меры информации.

Однако помимо общей судьбы, у каждого гена (то есть аллеля) есть и свои особенности, которые формируются под действием третьего фактора — отбора. Если бы не было отбора, то все гены давали бы чисто случайные отклонения от нейтрального уровня \bar{F}_{ST} .

Если же обнаружены достоверные отличия от селективно-нейтрального уровня, то мы можем выдвинуть гипотезу о действии на этот ген отбора.

Гены (аллели), изменчивость $F_{ST(i)}$ которых достоверно отличается от средней по всем локусам \bar{F}_{ST} , можно отнести к классу условно нейтральных генов ($F_{ST(i)} \approx \bar{F}_{ST}$).

Для генов, дифференциация которых достоверно ниже средней, можно предполагать действие стабилизирующего отбора ($F_{ST(i)} < \bar{F}_{ST}$).

В отношении генов, дифференциация которых достоверно выше средней, можно предполагать действие дифференцирующего отбора ($F_{ST(i)} > \bar{F}_{ST}$).

Такой анализ был проведен для всей ойкумены в целом и для отдельных регионов мира [6–11].

Показано, что в разных генофондах мира одни и те же гены подвержены разным типам отбора. Наименее устойчив класс селективно-нейтральных генов: показано, что ни один из генов не оказался селективно-нейтральным во всех регионах мира. В целом по структуре отбора генофонды регионов имеют незначительное сходство с общемировым генофондом (коэффициент корреляции $r \approx 0,3$). И лишь один регион — Северная Евразия (т. е. бывший СССР) занимает совершенно особое положение: уровень его сходства с глобальной селективной структурой ($r \approx 0,6$) в 2 раза выше среднерегионального! Причем высокие и достоверные корреляции в этом регионе обнаружены не только для селективной структуры в целом, но и для каждого ее класса. Для генов, подверженных стабилизирующему отбору: в Северной Азии $r_{STAB} \approx 0,6$ при среднерегиональной оценке $r_{STAB} \approx 0,3$; для условно нейтральных генов: $r_{NEUTR} \approx 0,5$ при средней $r_{NEUTR} \approx 0,15$; для генов, подверженных дифференцирующему отбору: в Северной Азии $r_{DIFF} \approx 0,7$ при средней $r_{DIFF} \approx 0,3$. Было обнаружено, что не только по структуре отбора, но по другим генетическим показателям генофонд Северной Евразии занимает особое, центральное место в структуре мирового генофонда [10, 11].

Итак, *селективная структура генофонда определяется типом и интенсивностью действия отбора на отдельные гены.*

Поскольку мы выдвигаем гипотезу отбора по наличию достоверных отклонений от среднего уровня дифференциации генов, то крайне важно провести эту операцию корректно. Поэтому мы разработали несколько вариантов технологии оценки отбора, причем каждая имеет свою сферу приложения [6–9].

Эти технологии имеют две важные особенности.

1. Выявлен истинный тип распределения оценок $F_{ST(i)}$ в генофонде ($F_{ST(i)}$ — дифференциация i -го аллеля в данном генофонде). Обнаружено [21], что из всех теоретических распределений (включая нормальное и логнормальное) только β -распределения дают хорошую аппроксимацию эмпирических распределений $F_{ST(i)}$, причем во всех генофондах мира и независимо от иерархического ранга региона. Поэтому мы оцениваем достоверность действия отбора согласно истинному β -распределению $F_{ST(i)}$.

2. Вторая особенность — моделирование выборочного процесса (бутстреппинг), которое проводится следующим образом. Из имеющегося набора генов (например, 93 полиморфных аллеля, изученных в населении Северной Евразии) случайным образом берется выборка по 50 генов, и попавшие в нее гены разносятся по классам селективной структуры (согласно β -распределению $F_{ST(i)}$). После того как такая процедура повторена миллион раз, определяется уточненная структура отбора: *гипотеза отбора выдвигается только в том случае, если в миллионе выборок ген всегда попал только в этот класс.*

Столь решительное требование позволяет отбирать те гены, для которых уверенно можно выдвигать гипотезу их подверженности отбору: они относятся классам стабилизи-

рующего отбора («LOWER DIFF»), дифференцирующего отбора («SUPER DIFF») или же селективной нейтральности («NEUTRAL»). Неустойчивые гены, которые в процессе моделирования попадали то в один, то в другой класс, помещаются в межклассовые области.

Таким сложным образом были изучены более 100 генных маркеров по всей ойкумене — в коренном населении каждой из частей света, а также Северной Евразии (бывшего СССР) и пяти ее субгенофондов: сибирского, среднеазиатского, уральского, кавказского, европейского [6–9].

При этом обнаружено, что несмотря на резкие различия между генофондами мира по структуре отбора, существует *общая закономерность*, которая выполняется во всех генофондах независимо от их иерархического ранга и состава населения. Эта закономерность позволила разработать процедуру *экспресс-диагностики отбора* [8]. Пользуясь ей, достаточно иметь только две оценки — $F_{ST(i)}$ для данного i -го гена и селективно-нейтральную оценку \bar{F}_{ST} — и можно без процедур аппроксимации и бутстреппинга ориентировочно *оценить тип и интенсивность отбора*.

Из серии полученных результатов укажем лишь один [9]. Он касается доли генов, подверженных отбору; в среднем эти гены составляют около 40%. Такое соотношение генов не зависит от иерархического ранга генофондов: для частей света доля селективно-значимых генов составила 38%, для субгенофондов Северной Евразии — 41%. В целом треть всех генов (32%) может быть уверенно отнесена к селективно-нейтральным, и 40% — к генам, подверженным действию отбора, из них 23% — давлению стабилизирующего отбора, 17% — дифференцирующего.

Важно отметить, что такая же оценка доли генов, подверженных отбору (40%), получена Боукок с коллегами [29], хотя они использовали совершенно иные методы оценки и анализировали ДНК-маркеры. По-видимому, полученное соотношение достаточно устойчиво, поскольку не зависит ни от метода, ни от типа маркеров, ни от ранга региона, ни от состава его населения.

Геногеография и отбор

Таким образом, имея два независимых метода изучения пространственной изменчивости генофонда — картографического моделирования и выявления структуры отбора можно проводить более объективный *комплексный* анализ генофонда. Рассмотрим комплексный анализ отбора на примерах двух генов.

Например, рассмотрим географию гена первичной гиполактазии (LAC^*R) в населении Старого Света [18]. Этот рецессивный ген в гомозиготе приводит к снижению активности лактазы у взрослых, что проявляется в непереносимости цельного молока. Существует гипотеза, что этот аллель был эволюционно исходным, и отбор шел в его пользу — непереносимость молока у взрослеющего ребенка приводила к его отказу от питания материнским молоком и повышала шансы на выживание следующего потомка. Распространение молочного животноводства выдвинуло противоположный отбор — преимущество получили индивиды, которые в любом возрасте усваивали цельное молоко. Предполагается, что именно этот фактор привел к резкому снижению частоты исходного гена.

Действительно, мы видим (рис. 1, а), что районы наиболее низких частот гена гиполактазии LAC^*R приходятся на побережья Северного и Балтийского морей; частота гена около 1%. У народов Урала его частота возрастает до 70–80% и далее на восток достигает 95%. Однако лишь на отдельных участках северной Европы наблюдаемый градиент коррелирует с интенсивностью развития молочного животноводства. Общая географическая

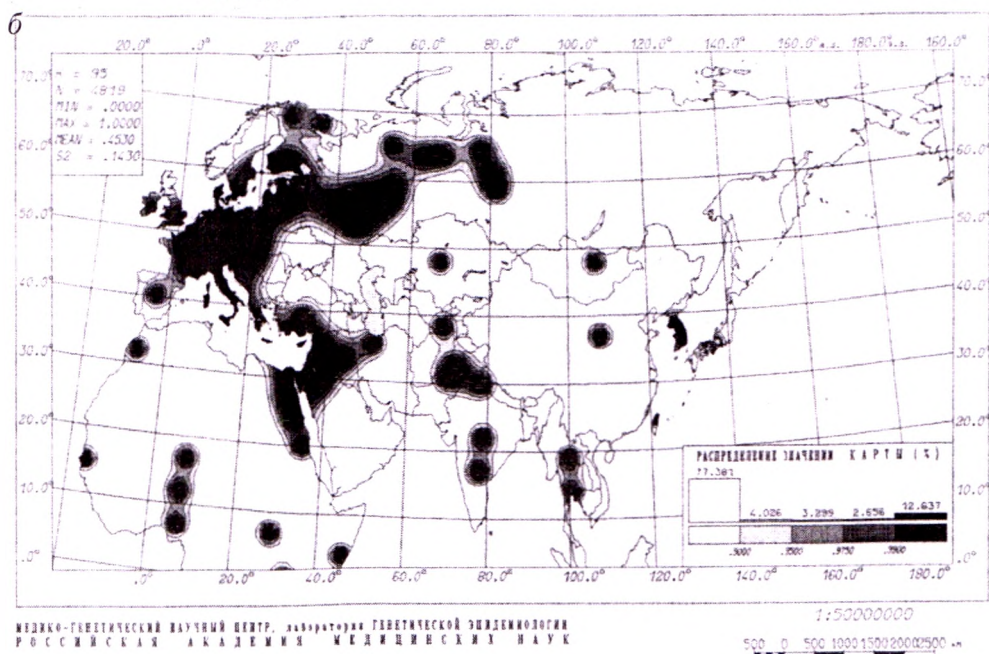
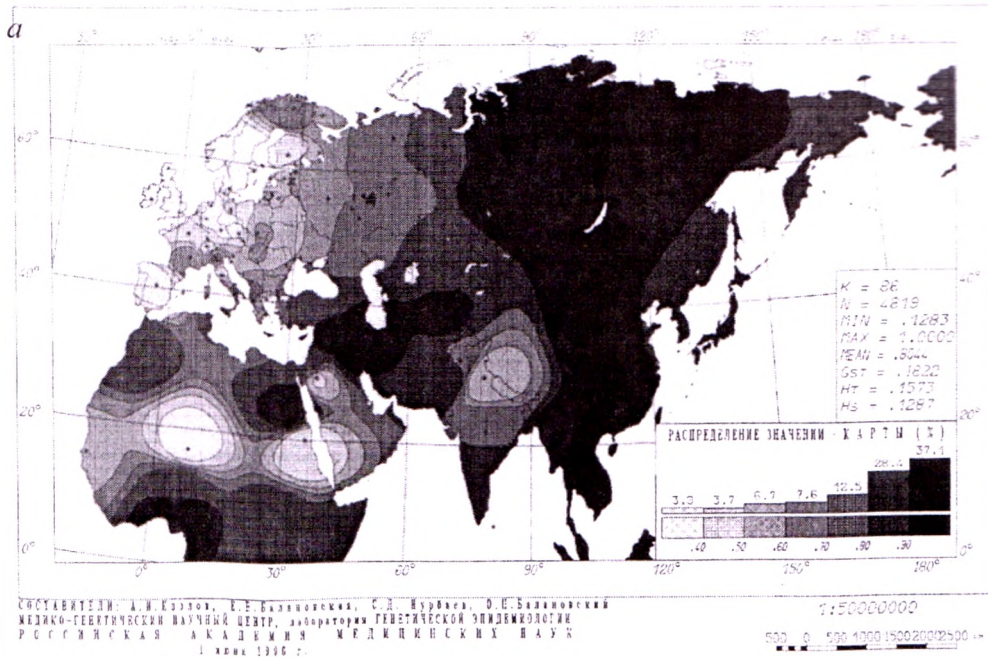


Рис. 1. Распространение частоты гена *LAC*R* (а) и надежность картографического прогноза (б) в населении Старого Света

закономерность, демонстрируемая картой, ставит под сомнение основную гипотезу — жесткой связи частоты этого гена с традицией молочного животноводства.

Карта выявляет еще одну особенность: дифференциация гена не только чрезвычайно велика, но для населения Европы она уникальна — $F_{ST(i)} \cdot 10^2 = 17$! Это значение в десять раз выше средней, селективно-нейтральной, изменчивости популяций этого региона. Ни по одному из генетических маркеров, изученных в народонаселении Европы (130 аллелей 35 локусов), не наблюдается столь высокой межпопуляционной дифференциации. Исходя из этого можно утверждать, что в Европе наблюдается мощный дифференцирующий отбор по гену первичной гиполактазии *LAC*R*: в селективной структуре генофонда Европы он занимает крайнее положение среди генов, подверженных дифференцирующему отбору.

Вопрос о природе столь мощного дифференцирующего отбора остается открытым. Чрезвычайно сомнительно, что особенности питания населения (напомним, что кисломолочные продукты полноценно усваиваются гомозиготами по *LAC*R*) могли стать столь мощным фактором и за столь малое эволюционное время привели к почти полному замещению гена *LAC*R* (если в Малой Азии традиции молочного животноводства возникли 6,3 тыс. лет до н. э., то в Скандинавии — 1,5 тыс. лет до н. э.)

С методической точки зрения анализ пространственной изменчивости *LAC*R* подчеркивает необходимость комплексного анализа при интерпретации результатов. Без анализа структуры отбора, сама геногеография гена не указывала на действие дифференцирующего отбора! Напротив, градиент частот гена почти полностью соответствует направлению селективно-нейтрального процесса — истории формирования населения этого региона.

Перейдем к другой карте (рис. 2, а) — гена кислого гликопротеина сыворотки крови человека (*AHSG*2*), география которого указывает на эффект отбора [32]. Корреляция частоты этого гена с географической широтой местности составила $r = 0,81$, а с интенсивностью ультрафиолетовой радиации (400–315 нм) $r = -0,82$. Таким образом, четко выраженная широтная изменчивость этого гена и его большая зависимость от климато-географических факторов прямо указывают на эффект отбора и адаптацию человека к климатическим условиям среды. Нам оставалось бы только добавить: на этот ген действует мощный дифференцирующий отбор — в северных широтах преимуществом преобладает этот аллель, а в южных — альтернативный. Однако анализ селективной структуры указывает, что этот вывод неверен: данный ген относится к генам, подверженным стабилизирующему отбору, а не дифференцирующему. Межпопуляционная изменчивость этого гена ($F_{ST(i)} \cdot 10^2 = 2,6$) не увеличена в сравнении с селективно-нейтральной ($F_{ST} \cdot 10^2 = 4,5$), а напротив — почти в два раза меньше! Таким образом, можно предположить, что дифференцирующий эффект климатических факторов подчинен более сильному действию стабилизирующего отбора по этому гену, по-видимому связанному с важной ролью гена в гомеостазе организма человека [32].

Рассмотрение этих двух карт указывает на необходимость комплексного подхода при изучении пространственной изменчивости: лишь в этом случае мы можем выйти из под власти той или иной схемы анализа и увидеть реальную жизнь гена с помощью ареальной генетики. Картографические технологии ареальной генетики создают целый спектр возможностей для количественного анализа пространственных закономерностей динамики гена в ареале популяции. Например, для экологической генетики чрезвычайно важно, что с помощью карты можно правильно оценить характер корреляции между признаками: карта позволяет видеть географию связи между признаками. При этом нередко наблюдается следующая картина: средний коэффициент корреляции $\bar{r} = 0$, при этом значительная часть карты демонстрирует корреляции, близкие к $r \approx -1$, а другая часть карты — к $r \approx +1$. Это означает, что корреляция между признаками чрезвычайно велика, но в разных группах

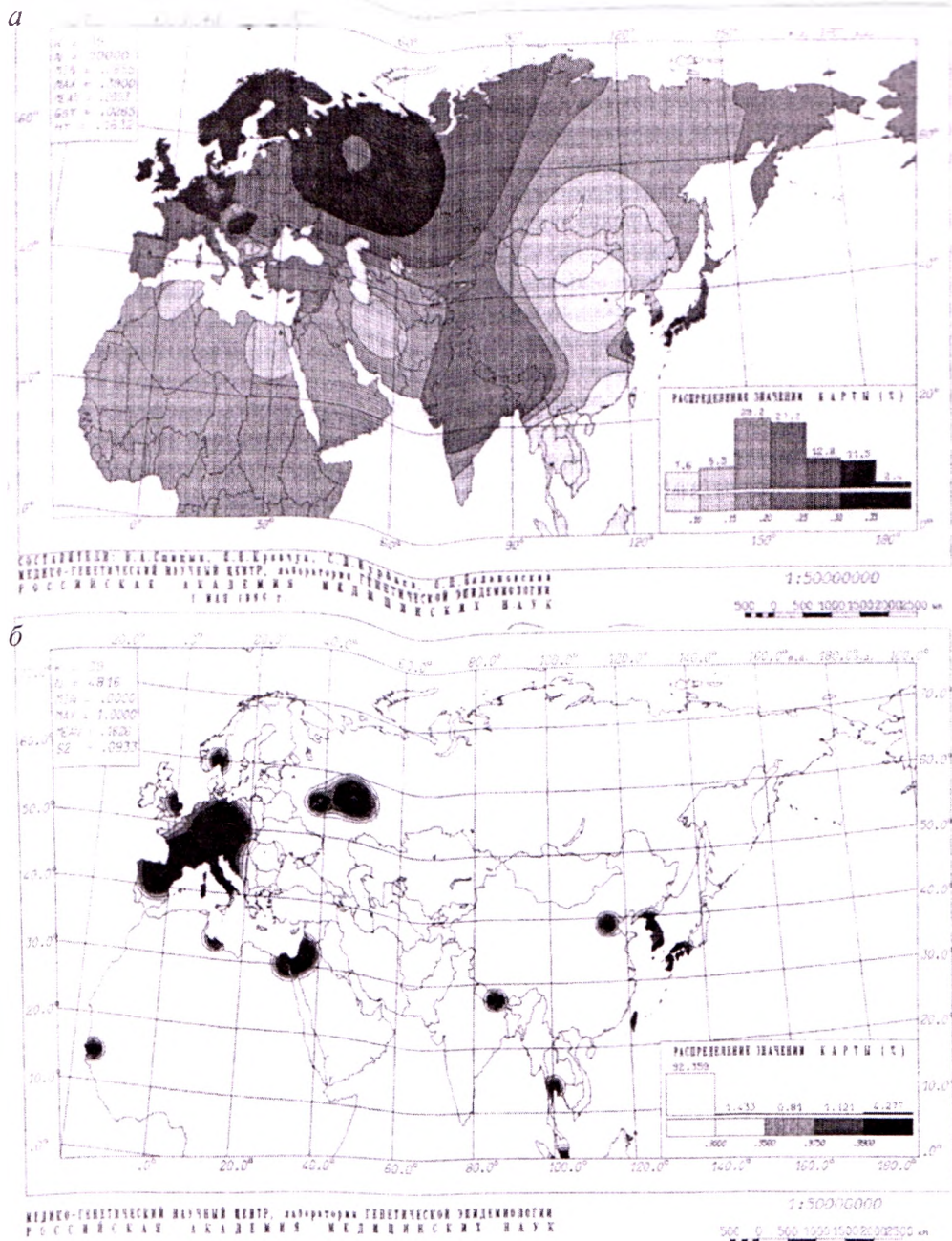


Рис. 2. Распространение частоты гена *AHS2* (а) и надежность картографического прогноза (б) в населении Старого Света

населения различается по направлению связи. Однако, если бы мы рассчитывали корреляцию без карты, мы пришли бы к неверному выводу, что связь вообще отсутствует ($r = 0$). Корреляционные карты позволяют не только правильно оценить связь между признаками, но выявить изменчивость этой связи в пространстве — районы, где она велика, или же те, где связь отсутствует. Этот и многие другие методы основаны на оригинальной технике «плывущего окна» [4].

Подчеркнем, что технология картографического моделирования применима не только на глобальном уровне, но и на уровне популяций любого ранга. Еще более важно то, что геногеографическая технология применима не только к полиморфным маркерам, но и к тем, которые находятся под таким полным контролем со стороны отбора, что их частоты близки к скорости мутаций. Например, при изучении коренного населения Марий Эл, было обнаружено, что синтетическая карта генов наследственной патологии [15] тесно коррелирует с синтетической картой классических генных маркеров [13]. Карта «нормального» генофонда мари демонстрирует те же пространственные закономерности, что и обобщенная карта генов наследственной патологии. Поскольку изменчивость генофонда в целом задается факторами миграций и дрейфа генов, то можно утверждать, что эти факторы определили географическую изменчивость не только классических генов, но и генов наследственной патологии. При этом уровень пространственной изменчивости генов наследственной патологии на несколько порядков ниже, чем нормальных генов: мощное давление отбора вызывает резкое снижение уровня дифференциации. Однако при этом характер изменчивости, его география остаются неизменными. Иными словами, если отбор действует независимо от географии, то даже гены, подверженные жесткому отбору, следуют тем же законам пространственной изменчивости, что и селективно-нейтральные гены, т. е. законам, определяемым факторами миграций и дрейфа генов. Более того, для населения Марий Эл показано, что карты отягощенности рецессивной и доминантной наследственной патологией практически идентичны — корреляция равна 0,9. Доминантная патология не может укрыться от отбора и ее частоты практически равны частоте мутаций. Таким образом, даже гены с таким низким уровнем полиморфизма (если, конечно, в анализ включается достаточно большой набор генов) могут анализироваться с помощью комплексной технологии, которую мы относим к области ареальной генетики: «ареальной» потому, что в основе ее лежит неперменный учет ареала популяции и что ее построения, как и любые модели, всегда стремятся к отражению реальности, но соответствуют ей лишь в недостижимом идеальном случае.

Надежность моделей

Именно поэтому в заключении обсудим самый важный вопрос — насколько можно верить картографическим моделям? Все мы, получая те или иные экспериментальные результаты, проводя их сравнения, всегда оцениваем степень достоверности. Обязателен вопрос, а какова она для карт? Ведь изучены далеко не все популяции в картируемом регионе, причем для разных маркеров набор изученных популяций обычно различен. Кроме того, как правило, проводится выборочное исследование, а достоверность выборочных оценок связана с размером выборки и частотой гена. Даже если и проведено тотальное обследование наследственной патологии, то и тогда оценка отягощенности зависит от общей численности населения района. Когда же мы переходим к генофондам высокого ранга (например, Старого Света), то этническая гетерогенность населения и гетерогенность при-

родной среды делают еще более сомнительными смелые выводы картографических построений.

Именно эти справедливые сомнения в достоверности карт тормозят развитие компьютерной картографии во всем мире [27, 31]. Однако попытки оценить достоверность карты привычными статистическими методами ни к чему не приводят — карта является столь сложной системой, что математический аппарат теории вероятности и математической статистики, на которых основана популяционная генетика, уже не позволяет найти решение проблемы.

Выход был найден несколько лет назад [19, 20], когда удалось ввести новый математический аппарат — теории надежности. Этот аппарат используется в технических науках для оценки надежности любых сложных систем: например, при 100%-ной надежности каждого отдельного узла, надежность системы в целом (ракеты, самолета, АЭС и т. д.) может быть невысока: она целиком зависит от характера системных связей (например, именно из-за низкой надежности системы в целом при 100%-ной надежности всех частей и была отвергнута программа СОИ). Геногеографические карты, конечно, проще, но ненамного. Поэтому теория надежности дает нам важный инструмент, позволяющий оценить надежность карты.

Впервые разработанные математические модели, алгоритмы и программное обеспечение позволяют оценивать надежность карты *в каждой ее точке* [19, 20]. Результатом стала новая карта, отражающая пространственную гетерогенность надежности. Карта надежности показывает: какова вероятность (P) того, что полученный результат совпадет с картографическим прогнозом, если провести генетическое исследование населения в любой точке карты. Это значение вероятности (P) и будет соответствовать оценке надежности картографической модели в данной точке пространства карты. Таким образом, надежность измеряется вероятностью осуществления прогноза (P) и оценивает статистическую достоверность каждого картированного значения.

Для простоты чтения карт надежности мы предложили объединить значения вероятности P в стандартные баллы, которые универсальны для любых карт. После долгих обсуждений с коллегами была принята шкала, наиболее приближающаяся к привычной пятибалльной системе и привычным уровням достоверности. «Единица» (1-й нурбалл: $0,00 \leq P \leq 0,90$), соответствующая на картах надежности «белым пятнам», указывает на самую низкую оценку надежности; «двойка» (2-й нурбалл: $0,90 < P \leq 0,95$) — несколько лучше, но надежность ниже 5%-ного уровня доверия; «тройка» (3-й нурбалл: $0,95 < P \leq 0,975$) — удовлетворительная оценка; «хорошо» (4-й нурбалл: $0,975 < P \leq 0,99$) — все еще ниже 1%-ного уровня доверия; пятый балл — «отлично» (5-й нурбалл: $0,99 < P \leq 1,00$) — соответствует наивысшей оценке надежности и отражается на карте надежности наиболее интенсивным цветом. Такую унифицированную систему оценок мы предложили назвать «нурбаллами».

Так, на приведенных картах (рис.1, б; 2, б) мы видим, что в картографо-статистический анализ включается далеко не вся площадь карты. Для той ее части, которая выглядит как белое пятно, также имеется картографический прогноз, но он не включается в анализ, поскольку не соответствует заданному уровню надежности (обычно уровню $P > 0,95$). Таким образом, при интерпретации карты можно выделить те области карты, которые определены надежно.

В настоящее время разработаны методы и программное обеспечение для учета трех важнейших факторов надежности.

1. Обеспеченность анализируемого узла сетки исходной генетической информацией, т. е. число и географическое положение изученных популяций.

2. Достоверность исходной информации, учитывающая размер выборки и частоту признака.

3. Гетерогенность ареала генофонда, приводящая к неравномерности потока генов: могут учитываться как социальные факторы (например, этнические или конфессиональные барьеры в распространении генов), так и природные (горные системы, акватории, пустыни и др.).

Важно подчеркнуть, что при проведении любых видов статистического анализа — главных компонент, корреляций и т. д., в анализе участвуют лишь те точки карты, которые определены как надежные. Таким образом мы получаем «надежную» геногеографию.

Эта работа является пионерской и имеет большое будущее, причем не только в геногеографии. Введение аппарата теории надежности позволяет решать многие трудные проблемы популяционной генетики. Укажем лишь две.

1. **Надежность редких признаков** (например, отягощенности наследственной патологией, частоты фамилий и т. д.). Если оценки получены тотально, т. е. обследована вся генеральная совокупность, то с точки зрения теории вероятности и математической статистики вопрос о достоверности оценок вообще не ставится. Однако ясно, что для выявления признаков с частотой одна миллионная генеральная совокупность должна быть не менее одного миллиона особей. Если, например, численность этнической группы всего 10 тыс. человек (а такова, например, численность адыгов-шапсугов) и мы сравниваем их с окружающим русским населением Краснодарского края (миллион человек), то ясно, что для оценки достоверности такого сравнения необходим специальный инструмент. Сейчас уже разработаны методы, математический аппарат и программное обеспечение для оценки надежности при сравнении популяций по отягощенности наследственной патологией (или по любым другим редким признакам). Разработаны рекомендации по планируемому объему выборок для выявления редких фенотипов с учетом спектра и частоты нозологических форм. Используя эти алгоритмы, мы можем оценивать достоверность оценок отягощенности даже при тотальном изучении популяции.

2. **Достоверность множественных сравнений (поправка Бонферрони).** При множественных сравнениях, например сравнении двух групп по полиаллельному локусу, чем больше аллелей мы тестируем, тем более строгий критерий достоверности различий для каждого аллеля мы обязаны ввести, чтобы на выходе получить все тот же 5%-ный критерий достоверности [см. 14, 17]. Теория надежности показывает, что стандартные требования (поправка Бонферрони) резко завышены: они соответствуют лишь редкому случаю. Разработанные алгоритмы и программы с использованием теории надежности позволяют оценить не единственный крайний случай, а все множество системных связей и найти оптимальные оценки статистической существенности при множественных тестах*.

* Представленные работы начинались в лаборатории генетики человека Института общей генетики РАН под рук. проф. Ю.Г.Рычкова. Основные результаты получены в лаборатории генетической эпидемиологии Медико-генетического научного центра РАМН под руководством чл.-кор. РАМН Е. К. Гинтера. Кроме участия этих коллективов (и, в первую очередь, их руководителей) представлены результаты совместной работы с В. А. Спицыным (и руководимым им коллективом лаборатории экологической генетики МГНЦ РАМН), с А. И. Козловым и В. П. Пасековым. Всем им, а также нашим многочисленным помощникам и коллегам мы приносим искреннюю благодарность.

УКАЗАТЕЛЬ ЛИТЕРАТУРЫ

1. Алтухов Ю. П. Внутривидовое генетическое разнообразие: мониторинг и принципы сохранения // Генетика. 1995. Т. 31, № 10. С. 1333–1357.
2. Балановская Е. В., Нурбаев С. Д., Рычков Ю. Г. Компьютерная технология геногеографического изучения генофонда. I. Статистическая информация карт // Генетика. 1994а. Т. 30, № 7. С. 951–965.
3. Балановская Е. В., Нурбаев С. Д., Рычков Ю. Г. Компьютерная технология геногеографического изучения генофонда. II. Статистическая трансформация карт // Генетика. 1994б. Т. 30, № 11. С. 1538–1555.
4. Балановская Е. В., Нурбаев С. Д. Компьютерная технология геногеографического изучения генофонда. III. Вычленение трендовых поверхностей // Генетика. 1995. Т. 31, № 4. С. 536–559.
5. Балановская Е. В., Нурбаев С. Д. Компьютерная технология геногеографического изучения генофонда. IV. Популяции в пространстве главных компонент // Генетика. 1997а. Т. 33, № 12. С. 1693–1710.
6. Балановская Е. В., Нурбаев С. Д. Селективная структура генофонда. I. Возможности изучения // Генетика. 1997б. Т. 33, № 11. С. 1574–1590.
7. Балановская Е. В., Нурбаев С. Д. Селективная структура генофонда. II. Стандартная технология определения через F_{ST} -статистики // Генетика. 1998а. Т. 34, № 9. С. 1307–1321.
8. Балановская Е. В., Нурбаев С. Д. Селективная структура генофонда. III. Технология определения через F_{ST} -статистики методом численного ресамплинга // Генетика. 1998б. Т. 34, № 10. С. 1434–1446.
9. Балановская Е. В., Нурбаев С. Д. Селективная структура генофонда. IV. Технология определения через показатель интенсивности отбора R_S // Генетика. 1998в. Т. 34, № 11. С. 1559–1573.
10. Балановская Е. В., Рычков Ю. Г. Этническая генетика: этногеографическое разнообразие генофонда народов мира // Генетика. 1990а. Т. 26, № 1. С. 114–121.
11. Балановская Е. В., Рычков Ю. Г. Этническая генетика. адаптивная структура генофонда народов мира по данным о полиморфных генетических маркерах человека // Генетика. 1990б. Т. 26, № 4. С. 739–748.
12. Балановская Е. В., Рычков Ю. Г. Генофонд человека на этапах освоения ойкумены: адаптивная эволюция и геногеография // Человек заселяет планету Земля. Глобальное расселение гоминид. М., 1997. С. 288–297.
13. Балановский О. П., Нурбаев С. Д., Кравчук О. И., Макаров С. В., Спицын В. А., Гинтер Е. К. «Синтетические» карты генофонда мари (по данным об иммуно-биохимическом полиморфизме) // Генетика. 1999. Т. 35, № 2. С. 312–321.
14. Вейр Б. Анализ генетических данных: дискретные генетические признаки. М., 1995. 400 с.
15. Гинтер Е. К., Мамедова Р. А., Козлова С. И., Галкина В. А., Мошкина И. С., Руденская Г. Е., Хлебникова О. В., Нурбаев С. Д., Балановская Е. В., Рассанов В. П. Разнообразие наследственной патологии в мариинской популяции и ее дифференциация по частотам генов наследственных болезней // Генетика. 1998. Т. 34, № 7. С. 963–972.
16. Грехова Л. В., Балановская Е. В., Рычков Ю. Г. Разработка технологии создания компьютерных региональных археологических атласов: поздний палеолит Северной Евразии // Гуманитарная наука в России: Сороковские лауреаты (История. Археология. Культурная антропология и этнография). М., 1996. С. 286–304.
17. Зайкин Д. В., Пудовкин А. И. Программа мульттест — расчет показателей статистической существенности при множественных тестах. // Генетика. 1991. Т. 27, № 11. С. 2034–2038.
18. Козлов А. И., Балановская Е. В., Нурбаев С. Д., Балановский О. П. Геногеография первичной гипоплазии в популяциях Старого Света // Генетика. 1998. Т. 34, № 5. С. 551–561.
19. Нурбаев С. Д., Балановская Е. В. Геногеография и генофонд. Оценивание надежности карты // Новые методы — новые подходы в современной антропологии. М., 1997. С. 116–132.
20. Нурбаев С. Д., Балановская Е. В. Компьютерная технология геногеографического изучения генофонда. V. Оценивание надежности карт // Генетика. 1998а. Т. 34, № 6. С. 825–838.
21. Нурбаев С. Д., Балановская Е. В. Межпопуляционное разнообразие генофонда: аппроксимация β -распределением // Генетика. 1998б. Т. 34, № 7. С. 837–841.
22. Реклю Э. Человек и Земля. Т. 1. СПб., 1906. Брокгауз и Эфрон.
23. Рычков Ю. Г. Генохронология исторических событий // Вопросы антропологии. 1986. Вып. 77. С. 3.
24. Рычков Ю. Г., Балановская Е. В. Обобщенный картографический анализ в антропологии. Отражение летописных славянских племен в антропологической географии современного русского населения // Вопросы антропологии. 1988. Вып. 80. С. 3–37.
25. Рычков Ю. Г., Балановская Е. В. Этническая генетика: соотношение адаптивной и нейтральной генетической дифференциации этносов // Генетика. 1990. Т. 26, № 3. С. 541–549.
26. Рычков Ю. Г., Балановская Е. В. Геногеография народонаселения СССР // Генетика. 1992. Т. 28, № 1. С. 52–75.

27. *Barbujani G., Nasidze I. S., Whitehead G. N.* Genetic diversity in the Caucasus // *Human Biol.* 1994. Vol. 66, № 4. P. 639–668.
28. *Bowcock A., Cavalli-Sforza L.* The study of variation in the human genome // *Genomic.* 1991. Vol. 11. P. 401–498.
29. *Bowcock A. M., Kidd J. R., Mountain J. L., Hebert J. M., Carotenuto L., Kidd K. K., Cavalli-Sforza L. L.* Drift, admixture, and selection in human evolution: A study with DNA polymorphisms // *Proc. Natl. Acad. Sci. USA.* 1991. Vol. 88. P. 839–843.
30. *Cavalli-Sforza L. L., Bodmer W. F.* The Genetics of Human Population. San Francisco, 1971. 965 p.
31. *Cavalli-Sforza L. L., Menozzi P., Piazza A.* History and Geography of Human Genes. Princeton: 1994. 1066 p.
32. *Spitsyn V. A., Kravchuk O. I., Nurbaev S. D., Krause D., Kuchheuser W.* Climate-dependent genetic variation of alpha-2HS-glycoprotein // *Human Biology.* 1998. Vol. 70, №3. P. 463–475.